

---

## Distributed Computing on HTCondor - Practical

### 1 Introduction

Welcome to the practical session for the Distributed Algorithms tutorial on **Distributed Computing on the University of Liverpool's HTCondor Pool**. This session will focus on utilising the theoretical aspects discussed throughout the lecture in practice. If you have any questions, please send an email to **matthew.carter@liverpool.ac.uk**.

**NOTE: You will not receive feedback on the work completed throughout this practical session. If you do not finish all parts of the practical session, you should endeavour to complete them in your spare time.**

### 2 Topics Covered

Throughout this practical session, you will:

- Access the Universities High Throughput Computing (HTCondor) Service
- Edit and write a series of programs on HTCondor
- Submit jobs to the scheduler and run them on the pool
- Design and implement your own embarrassingly parallel algorithm

### 3 Accessing HTCondor

Firstly, there are multiple ways of logging into HTCondor. As long as you are connected to the University's network or the University's VPN, you can remotely access HTCondor using any Command Line Interface.

These labs will assume you are on a University computer and are using MobaXterm. Follow these steps to gain access to HTCondor.

- Find 'MobaXterm' by using Window's search icon. (If it is not there, you can use the app 'Install University Applications' to search for and install 'MobaXterm'.)
- Click on 'Session' in the top left.
- Click on 'SSH'.
- In the 'Remote host' field, type in "condor.liv.ac.uk".

- Make sure that 'Specify username' is enabled.
- In the 'Specify username' field, type in your MWS username (what you use to login to the University computers).
- Finally, click 'OK', and follow the instructions in the Command Line Interface.
- You MUST work from your data storage area on HTCondor, located at '/condor\_data/mwsusername'.

**NOTE: Using a command line interface to connect to the Universities HTCondor pool forces you to use a command line text editor such as nano. This is extremely tedious and leads to mental pain. For your sanity, I recommend connecting through VSCode (described below) or PyCharm.**

To access HTCondor from home, you must request access to the VPN services. To do this, follow these steps:

- Go to 'servicedesk.liverpool.ac.uk'.
- Login with your MWS username and password.
- Click 'Request'.
- Click 'Register for Virtual Private Network'.

You can also connect to the Universities HTCondor pool via VSCode. To do this, follow these steps:

- Open VS Code, and go to 'extensions'.
- Search for Remote - SSH, and install it.
- Press 'f1' and type in "Remote-SSH: Add new host"
- Type in "mwsusername@condor.liv.ac.uk"
- The host should have been added so press 'f1' again.
- Type in "Remote-SSH: Connect to Host"
- When prompted, click "Linux" and that you trust the fingerprint.
- Click the 'Explorer' in the top left and click 'Open Folder'
- You can ensure that you work from your data area on HTCondor, which is at '/condor\_data/mwsusername'. Click 'ok'

- Type in your password again, and you should connect.
- Press **Ctrl+Shift+’** and you should have a terminal and can open files from HTCondor on VS Code

## 4 Hello World

Throughout this section, we will run a simple *Hello World* program on the HTCondor pool which performs a Monte Carlo simulation to estimate the quantity of Pi. To run this program on HTCondor, you should:

1. Ensure that you are in your user area by typing:  
**cd /condor\_data/mwsusername**
2. Clone the Github repository by typing:  
**git clone https://github.com/mjcarter95/UoL-HTCondor-101**
3. Change into the directory containing the program:  
**cd ./UoL-HTCondor-101/python\_applications**
4. Submit the job file to the scheduler:  
**condor\_submit python\_submit.sh**
5. Wait for the jobs to finish, you can manage their status with **condor\_q**
6. Process the results by typing:  
**python3 output.py**

The file **python\_submit.sh** contains the instructions for requesting resources for computation and running the program on the HTCondor pool. Before moving to the next section, it would help to familiarise yourself with how this HTCondor program works.

## 5 Search for a Word in the Dictionary

Design a distributed algorithm to search for a specific word in a dictionary dataset distributed across multiple nodes. You can download a text file of  $\sim 500,000$  English words from <https://github.com/dwyl/english-words>. You should:

- Write a Python program to process the dictionary into N datasets
- Write a Python program to search for a word within a dataset

- Write a Slurm job description to run the Python program on N nodes in the HTCCondor pool
- Write a Python file to process the output

## 6 End of Session

That's all for today's practical session! Today you have learnt:

- How to login to HTCCondor.
- How to view, and edit code.
- How to compile and run programs
- Submitting a job to the compute node
- Designed and implemented an embarrassingly parallel algorithm